*Environmental Health Perspectives*
*Vol. 32, pp. 83-101, 1979*

# Multivariate Analysis of Cancer Mortalities for Selected Sites in 24 Countries

by Haruo Yanai,* Yutaka Inaba,† Hirohumi Takagi,‡ and Shunichi Yamamoto†

**In order to analyze the pattern in the geographical distribution of cancer death in 24 countries of the world, correlation coefficients were calculated between pairs of mortality rates of different cancer sites, using the data for 13 sites in males and 14 sites in females over 18 years from 1950 to 1967. Then factor analysis by means of varimax method was performed on 13 × 13 correlation matrix for males, 14 × 14 correlation matrix for females and 27 × 27 correlation matrix for males and females combined.**

**As a result of factor analysis, three factors are extracted, which are commonly recognized in both males and females. The first factor has high positive factor loadings on pancreas, prostate (for males), skin, and intestine cancers, and negative loadings on stomach and liver cancers. The second factor has high positive factor loadings on rectum, intestine, and lung cancers, and the third factor on larynx, oral, and esophagus cancers. Factor analysis based on 27 × 27 correlation matrix revealed that the third factor of both sexes are heterogeneous with regards to the distributions of the factor score.**

**In order that we may find some clues to develop an etiological hypothesis for each site of cancer, we obtained the correlation coefficient between the scores of the extracted factors and the variables on food and environmental agent, and performed stepwise regression methods as well. One of the most striking results we obtained was that excessive drinking of alcohol and the lack of appropriate intake of fruit are suspected as etiological promoters in the pathogenesis of oral, esophagus, and larynx cancers in males.**

## Introduction

So far, the spatial distribution of cancer mortality and morbidity rates has been studied by many authors, and its characteristic pattern as to the various sites of cancer has stimulated much concern for the biostatisticians in various countries. Burkit (*1, 2*) stated that the variation in the geographical pattern of a disease may be related to its cause, and in view of this assertion many authors have studied the geographic distribution of cancer.

Segi (*3-5*) has published a number of elaborate works on cancer mortality of various sites in some countries including Japan. Dunham et al. (*6*) mapped mortality rates of 18 cancer sites for about 100 geographical areas in the world. Hoover (*7*) analyzed the geographic pattern of cancer mortality in the United States from 1950 to 1969, and Waterhouse (*8*) reported the demographical pattern of cancer incidence in U.K.

In order to examine Burkit's assumption that the similar geographic distribution of two different cancers may suggest the existence of a common etiological cause between them, it is necessary to calculate correlation coefficients between pairs of the various cancer sites with various sites. As it has been generally recognized that any of the specific factors do not contribute independently but interdependently to the etiology in chronic diseases including cancer, some methods of multivariate analysis may be used to analyze the complicated structural interdependence among the many correlated variables. Among the available methods of multivariate analysis, factor analysis is quite useful as a first step to uncover the unknown etiology which may be hidden behind the data.

We have already completed work (*9*) which

*Department of Psychology, Chiba University, Yayoicho, Chiba City, Japan. Author to whom reprint requests should be sent.

†Department of Hygiene, Faculty of Medicine, University of Tokyo, Tokyo, Japan.

‡Department of Epidemiology, Faculty of Medicine, University of Tokyo, Tokyo, Japan.

analyzed the interrelationship between the mortality rate of any cancer sites in terms of the geographic and age distributions in Japanese cases and extracted four factors, which are commonly recognized in both sexes. In this paper, we attempt to analyze the characteristic pattern of geographic distribution of mortality rates of various sites of cancer in the world, by using the series of data collected by Segi and Kurihara (*4*).

# Method

## Data

The data used in the analysis were obtained from age-adjusted death rates for malignant neoplasms for selected sites in 24 countries from 1950 to 1967 (*4*). The list of 24 countries is shown in Table 1. The number of sites of cancer selected for analysis is 13 for males and 14 for females, as shown in Table 2. Note that age-adjusted death rates for these cancer sites were calculated by using standard population of 46 countries around 1950. In addition to the data of cancer mortality, we used the data from FAO Production Year Book (*10*), in which calories per capita per day with respect to various kinds of food such as milk, meat, or oil and fats are described. We also used the data concerning environmental variables such as population density, rainfall, and so on. The list of these variables is shown in Table 3. The missing data were replaced by the mean mortality rates calculated from available data.

## Procedures of Analysis

***Correlation coefficients on the geographical distribution.*** Since the age-adjusted mortality rates of 24 countries for various cancer sites are chronologically given as the mean values of the two consecutive calendar years; (1) 1950-51, (2) 1952-53, (3) 1954-55, (4) 1956-57, (5) 1958-59, (6) 1960-61, (7) 1962-63, (8) 1964-65, (9) 1966-67, these nine periods are classified into the following three categories: (1) 1950-55, (2) 1956-61, and (3) 1962-67, and the mean values for the three chronological categories are computed. Using these values, we calculated correlation coefficients between each pair of the three periods for each site and sex. As shown in Tables 5 and 6 below, fairly high correlation coefficients were found for most of the cancer sites. We thus calculated the means of the age-adjusted mortality rates of these three periods for each country. We call them simply mortality rate hereafter instead of the mean age-adjusted mortality rates over 18 years. Next, we calculated correlation coefficients between the male and female mortality rates for each cancer site excluding prostate, breast and uterus cancers. Finally, we calculated correlation coefficients between pairs of site-specific mortality rates of 13 sites for males and 14 sites for females. Hence, we have a $13 \times 13$ correlation matrix for males and a $14 \times 14$ correlation matrix for females.

***Factor Analysis.*** Factor analysis (*11*) was performed for the $13 \times 13$ correlation matrix (for males), the $14 \times 14$ correlation matrix (for females) and the $27 \times 27$ correlation matrix (for males and females combined) respectively. The procedure of factor analysis adopted in this study was as follows: First, communality for each variable was estimated by squared multiple correlation (SMC) by Guttman (*12*); principal factor loadings are obtained by Thomson's refactoring method, and factor loading matrix is rotated by Kaiser's Varimax method (*13*) in order to obtain a simple structure solution. Next, we estimate a factor score matrix $\hat{F}$ by $\hat{F} = ZR^{-1}A$, where $A$ is a factor loading matrix. Finally, we calculate correlation coefficients between each factor score and 24 variables shown in Table 3.

**Table 1. List of 24 countries.**

| No. | Country |
|---|---|
| ( 1) | Union of South Africa[a] |
| ( 2) | Canada[b] |
| ( 3) | Chile |
| ( 4a) | United States, white [c] |
| ( 4b) | United States, nonwhite[c] |
| ( 5) | Israel[d] |
| ( 6) | Japan |
| ( 7) | Germany Federal Republic[e] |
| ( 8) | Austria |
| ( 9) | Belgium |
| (10) | Denmark |
| (11) | Finland |
| (12) | France |
| (13) | Ireland |
| (14) | Italy |
| (15) | Norway |
| (16) | Netherland |
| (17) | Portugal |
| (18) | England and Wales |
| (19) | Scotland |
| (20) | Northern Ireland |
| (21) | Sweden |
| (22) | Switzerland |
| (23) | Australia[f] |
| (24) | New Zealand[g] |

[a]Including the European population only.
[b]Excluding the Yukon and the North-West territories from 1950 to 1957, including these areas from 1958 to 1967.
[c]Excluding New Jersey for 1962-1963.
[d]Including the Jewish population only.
[e]Excluding West Berlin from 1950 to 1961 and including West Berlin from 1962 to 1967.
[f]Excluding full-blood aboriginals.
[g]Excluding Maoris from 1950 to 1961, including them from 1962 to 1967.

Table 2. List of the sites used in the analysis.

| I.C.D. Code number | Site | Abbreviation for analysis | | |
|---|---|---|---|---|
| 140-148 | Oral cavity and pharynx | Oral | male | female |
| 150 | Esophagus | Esophagus | male | female |
| 151 | Stomach | Stomach | male | female |
| 152 | Intestine (small and large) | Intestine | male | female |
| 154 | Rectum | Rectum | male | female |
| 155 | Liver, biliary passage | Liver | male | female |
| 157 | Pancreas | Pancreas | male | female |
| 161 | Larynx | Larynx | male | female |
| 162 | Lung, bronchus, trachea | Lung | male | female |
| 170 | Breast | Breast | | female |
| 171 | Cervix uteri, corpus uteri | Uterus | | female |
| 177 | Prostate | Prostate | male | |
| 190-191 | Skin | Skin | male | female |
| 194 | Thyroid gland | Thyroid | male | female |
| 204 | Leukemia, aleukemia | Leukemia | male | female |

Table 3. Foods and environmental variables studied.

Foods
- Alcohol[a]
- Energy (total intake)
- Vegetable products
- Animal products
- Cereals
- Roots and tubers
- Sugar
- Pulses (beans, peas)
- Tree nuts
- Vegetables
- Fruits
- Meat
- Eggs
- Fish
- Milk
- Oil fats
- Vegetable oil
- Animal oil
- Spices
- Stimulants (tea, coffee)[a]

Environmental variables
- Rate of workers in primary industries (Primary industry)
- Population density
- Mean temperature in metropolis in 1965
- Rainfall in metropolis in 1965

[a]UN data (*10*).

***Stepwise Regression Method.*** In order to examine the causal relationship between cancer mortality and food or environmental variables, the stepwise forward regression method (*14*) was performed, using each of the cancer mortality rates of various sites as a criterion, and 16 variables were selected out of the 24 variables in Table 3 as a set of explanatory variables.

# Results

## Site-Specific Mortality Rates in Each Period

The site-specific mortality rates of cancer in each of three periods are shown in Table 4 and in its last column sex ratio of cancer mortality is given. In the first period, the mortality rate of stomach cancer ranks highest for both sexes, but it decreases as the calendar year proceeds, and the mortalities of lung cancer increase instead. Furthermore, the mortality rate of breast cancer increases and that of uterus cancer decreases as the year proceeds. As for the sex ratio, larynx cancer ranks highest, followed by lung, oral, and esophagus cancers.

## Correlation Coefficients with Regard to the Geographical Distributions of Cancer Mortality

We have shown the correlation coefficient of the mortality rates for each pair of three periods: (1) from 1950 to 1955, (2) from 1956 to 1961, and (3) from 1962 to 1967 in Table 5 for males and in Table 6 for females. It is seen from these tables that most of the sites have consistently high correlation with few exceptions. Such consistently high correlations indicate that the chronological variation in the geographical distributions of various cancer sites is stable throughout the whole period from 1950 to 1967. Of all the geographical distributions concerning the mortalities of cancer sites, that of liver cancer is the most stable throughout these periods in both sexes, followed by intestine cancer in both sexes and breast cancer in females, while those of pancreas cancer in both sexes and that of oral cancer in males are relatively unstable. Next, we show in Table 7 the correlation coefficients between the male and female

**Table 4. Site-specific mortality rate (per 100,000) with respect to sex and three time periods.**

| Site | Period[a] | Mortality rate/100,000 Male | Female | Ratio male/female |
|---|---|---|---|---|
| Oral | (1) | 3.952 | 1.231 | 3.210 |
| | (2) | 3.785 | 1.174 | 3.224 |
| | (3) | 3.649 | 1.141 | 3.198 |
| Esophagus | (1) | 5.213 | 2.016 | 2.586 |
| | (2) | 5.047 | 1.873 | 2.695 |
| | (3) | 5.042 | 1.857 | 2.715 |
| Stomach | (1) | 36.130 | 22.043 | 1.639 |
| | (2) | 32.511 | 18.764 | 1.733 |
| | (3) | 28.240 | 15.539 | 1.817 |
| Intestine | (1) | 9.908 | 10.317 | 0.960 |
| | (2) | 10.005 | 10.320 | 0.969 |
| | (3) | 10.398 | 10.382 | 1.002 |
| Rectum | (1) | 7.069 | 4.328 | 1.633 |
| | (2) | 6.719 | 4.196 | 1.601 |
| | (3) | 6.517 | 4.153 | 1.569 |
| Liver | (1) | 6.447 | 7.111 | 0.907 |
| | (2) | 6.151 | 6.336 | 0.971 |
| | (3) | 6.067 | 5.804 | 1.045 |
| Pancreas | (1) | 4.869 | 3.361 | 1.449 |
| | (2) | 6.066 | 3.887 | 1.561 |
| | (3) | 6.918 | 4.343 | 1.593 |
| Larynx | (1) | 2.438 | 0.362 | 6.735 |
| | (2) | 2.481 | 0.317 | 7.826 |
| | (3) | 2.668 | 0.291 | 9.168 |
| Lung | (1) | 20.514 | 3.904 | 5.255 |
| | (2) | 38.383 | 4.516 | 8.499 |
| | (3) | 35.350 | 5.311 | 6.656 |
| Skin | (1) | 1.817 | 1.354 | 1.342 |
| | (2) | 1.920 | 1.420 | 1.352 |
| | (3) | 1.998 | 1.448 | 1.380 |
| Thyroid | (1) | 0.483 | 0.885 | 0.546 |
| | (2) | 0.491 | 0.883 | 0.556 |
| | (3) | 0.469 | 0.818 | 0.573 |
| Leukemia | (1) | 4.905 | 3.673 | 1.335 |
| | (2) | 5.864 | 4.219 | 1.390 |
| | (3) | 6.198 | 4.403 | 1.408 |
| Prostate | (1) | 10.988 | — | — |
| | (2) | 12.416 | — | — |
| | (3) | 13.087 | — | — |
| Breast | (1) | — | 17.528 | — |
| | (2) | — | 18.201 | — |
| | (3) | — | 18.919 | — |
| Uterus | (1) | — | 14.775 | — |
| | (2) | — | 13.392 | — |
| | (3) | — | 11.965 | — |

[a](1) First period, from 1950 to 1955; (2) second period, from 1956 to 1961; (3) third period, from 1962 to 1967.

mortality rates for 12 cancer sites, using their mean values over 18 years. Low correlation coefficients are observed in oral, esophagus, and larynx cancers, and moderate degrees of correlation are found in lung and liver cancers. The remaining sites of cancer have considerably high correlation coefficients.

We calculated the correlation coefficients between any pair of the sites with regard to the geographical distribution. The 13 × 13 correlation ma-

**Table 5. Correlation coefficients between the mortalities of any two of two of the three periods for each site of male cancer.**

| Site | Period 1-2[a] | Period 1-3[a] | Period 2-3[a] |
|---|---|---|---|
| Oral | 0.886 | 0.688 | 0.938 |
| Esophagus | 0.936 | 0.818 | 0.959 |
| Stomach | 0.967 | 0.916 | 0.982 |
| Intestine | 0.969 | 0.932 | 0.975 |
| Rectum | 0.978 | 0.947 | 0.982 |
| Liver | 0.986 | 0.966 | 0.987 |
| Pancreas | 0.897 | 0.853 | 0.957 |
| Larynx | 0.976 | 0.946 | 0.989 |
| Lung | 0.987 | 0.949 | 0.985 |
| Skin | 0.954 | 0.917 | 0.970 |
| Thyroid | 0.952 | 0.939 | 0.956 |
| Leukemia | 0.952 | 0.893 | 0.957 |
| Prostate | 0.962 | 0.924 | 0.977 |

[a]As in Table 4.

**Table 6. Correlation coefficients between the mortalities of any two of the three periods for each site of female cancer.**

| Site | Period 1-2[a] | Period 1-3[a] | Period 2-3[a] |
|---|---|---|---|
| Oral | 0.967 | 0.902 | 0.940 |
| Esophagus | 0.965 | 0.881 | 0.955 |
| Stomach | 0.972 | 0.906 | 0.976 |
| Intestine | 0.983 | 0.953 | 0.977 |
| Rectum | 0.963 | 0.893 | 0.944 |
| Liver | 0.966 | 0.946 | 0.985 |
| Pancreas | 0.871 | 0.745 | 0.923 |
| Larynx | 0.899 | 0.856 | 0.964 |
| Lung | 0.932 | 0.847 | 0.934 |
| Skin | 0.947 | 0.847 | 0.935 |
| Thyroid | 0.959 | 0.934 | 0.964 |
| Leukemia | 0.947 | 0.889 | 0.954 |
| Breast | 0.986 | 0.974 | 0.991 |
| Uterus | 0.969 | 0.929 | 0.984 |

[a]As in Table 4.

**Table 7. Correlation coefficients between the male mortality rate and the female mortality rate for 12 sites of cancer with regard to the geographical distribution.**

| Site | Correlation coefficient |
|---|---|
| Larynx | 0.368 |
| Esophagus | 0.300 |
| Stomach | 0.976 |
| Intestine | 0.966 |
| Rectum | 0.956 |
| Liver | 0.780 |
| Pancreas | 0.923 |
| Larynx | 0.170 |
| Lung | 0.620 |
| Skin | 0.940 |
| Thyroid | 0.894 |
| Leukemia | 0.960 |

trix in males and the 14 × 14 correlation matrix in females are shown in Tables 8 and 9, respectively.

By convention of inferential statistics, correlation coefficients with the value of 0.394 or more were considered statistically significant at the 5% level, based upon the assumption that the mortality rates computed were random samples from the population characterized by normal distribution. The number of pairs with statistically significant correlation at the 5% level was 28 (32.1%) in males and 31 (34.1%) in females. The highest correlation coefficient was found between rectum and intestine cancers in males ($r = 0.754$) and between breast and intestine cancers in females ($r = 0.817$), while the highest negative correlation coefficient was observed between stomach and intestine cancers in both sexes (for males $r = -0.729$, and for females, $r = -0.733$).

## Factor Analysis

First, we performed a series of steps required for factor analysis procedure on the 13 × 13 and 14 × 14 correlation matrices. Results are shown in Tables 8 and 9, respectively. We extracted four factors by this procedure for each sex. The contribution of these factors totaled 65.5% and 65.3% of the total in males and females, respectively.

The factor loading matrices of males and females are shown in Tables 10 and 12 respectively. The scores of the corresponding four factors with respect to 24 countries are obtained and shown in Table 11 for males and Table 13 for females. Note that we exchanged the order of the first and second factors in females so that they correspond to those in males in terms of the order in which the factors $f_j$ are written. Figures 1 and 3 graphically display the factor loadings of the first (vertical line) and second factors (horizontal line) in males and females respectively, and Figures 2 and 4 show those of the third (vertical line) and fourth factors (horizontal line). In addition, we drew a configuration of factor scores of 24 countries in Figures 5-8 corresponding to the figures of the factor loadings as shown in Figures 1-4. Now, let us inspect in more details what patterns the four factors display in view of Tables 10 to 13, and Figure 1-8.

***First Factor.*** In males, high factor loadings were found for pancreas, prostate, intestine, and leukemia, (positive in sign) and liver (negative in sign),

**Table 8. Correlation coefficients between the mortalities of any pair of 13 sites of male cancer with regards to geographical distribution.**

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( 1) | Oral | 1.000 | | | | | | | | | | | | |
| ( 2) | Esophagus | 0.457 | 1.000 | | | | | | | | | | | |
| ( 3) | Stomach | −0.404 | 0.344 | 1.000 | | | | | | | | | | |
| ( 4) | Intestine | 0.451 | −0.221 | −0.729 | 1.000 | | | | | | | | | |
| ( 5) | Rectum | 0.100 | −0.175 | −0.339 | 0.754 | 1.000 | | | | | | | | |
| ( 6) | Liver | −0.149 | 0.317 | 0.567 | −0.553 | −0.266 | 1.000 | | | | | | | |
| ( 7) | Pancreas | 0.220 | −0.166 | −0.562 | 0.552 | 0.158 | −0.724 | 1.000 | | | | | | |
| ( 8) | Larynx | 0.598 | 0.582 | −0.056 | −0.018 | −0.063 | 0.421 | −0.235 | 1.000 | | | | | |
| ( 9) | Lung | 0.130 | −0.001 | −0.107 | 0.516 | 0.589 | −0.243 | 0.309 | −0.020 | 1.000 | | | | |
| (10) | Skin | 0.386 | −0.188 | −0.429 | 0.462 | 0.097 | −0.355 | 0.336 | −0.001 | 0.114 | 1.000 | | | |
| (11) | Thyroid | 0.125 | 0.289 | 0.155 | −0.019 | 0.128 | 0.015 | −0.005 | −0.077 | 0.162 | 0.042 | 1.000 | | |
| (12) | Leukemia | −0.035 | −0.387 | −0.603 | 0.423 | 0.329 | −0.386 | 0.395 | −0.166 | 0.138 | 0.392 | 0.172 | 1.000 | |
| (13) | Prostate | 0.279 | −0.040 | −0.575 | 0.516 | 0.247 | −0.484 | 0.649 | −0.043 | 0.218 | 0.372 | 0.157 | 0.543 | 1.000 |

**Table 9. Correlation coefficients between the mortalities of any pair of 14 sites of female cancer with regards to geographical distribution.**

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( 1) | Oral | 1.000 | | | | | | | | | | | | | |
| ( 2) | Esophagus | 0.374 | 1.000 | | | | | | | | | | | | |
| ( 3) | Stomach | −0.309 | 0.518 | 1.000 | | | | | | | | | | | |
| ( 4) | Intestine | 0.486 | −0.218 | −0.733 | 1.000 | | | | | | | | | | |
| ( 5) | Rectum | 0.237 | −0.135 | −0.300 | 0.675 | 1.000 | | | | | | | | | |
| ( 6) | Liver | −0.771 | −0.222 | 0.407 | −0.523 | −0.164 | 1.000 | | | | | | | | |
| ( 7) | Pancreas | 0.543 | 0.174 | −0.438 | 0.540 | 0.238 | −0.489 | 1.000 | | | | | | | |
| ( 8) | Larynx | 0.566 | 0.459 | 0.044 | 0.219 | 0.077 | −0.216 | 0.085 | 1.000 | | | | | | |
| ( 9) | Lung | 0.413 | 0.333 | −0.096 | 0.399 | 0.397 | −0.153 | 0.550 | 0.356 | 1.000 | | | | | |
| (10) | Skin | 0.181 | −0.203 | −0.457 | 0.481 | 0.178 | −0.278 | 0.236 | 0.042 | 0.108 | 1.000 | | | | |
| (11) | Thyroid | 0.048 | 0.056 | 0.168 | −0.118 | 0.029 | 0.148 | 0.299 | −0.077 | 0.220 | 0.221 | 1.000 | | | |
| (12) | Leukemia | −0.095 | −0.400 | −0.530 | 0.287 | 0.244 | 0.004 | 0.326 | −0.591 | −0.004 | 0.440 | 0.237 | 1.000 | | |
| (13) | Breast | 0.342 | −0.287 | −0.724 | 0.817 | 0.627 | −0.338 | 0.573 | −0.077 | 0.410 | 0.495 | 0.149 | 0.628 | 1.000 | |
| (14) | Uterus | −0.253 | −0.022 | 0.291 | −0.228 | −0.018 | 0.167 | 0.028 | −0.156 | −0.027 | −0.440 | −0.046 | −0.390 | −0.298 | 1.000 |

producing the high degree of contribution ratio (27.9%). On the other hand, in females, high loadings were found only for skin, breast, stomach (negative), and uterus (negative) cancers and leukemia, resulting in a relatively smaller degree of the contribution ratio as compared to the male case. Hence, the first factor of males does not correspond to that of females in the strict sense but is sure to reflect some common factors. With regard to the scores of the first factor in males shown along with the vertical line in Figure 1, U.S. white is plotted highest, followed by U.S. nonwhite and Australia; Japan is plotted lowest, while among females Norway is plotted highest, and Japan and Chile lowest.

**Table 10. Factor loadings of 13 sites of cancer for males.**

| Sites | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Communality |
|---|---|---|---|---|---|
| Oral | 0.344 | 0.081 | 0.793 | 0.169 | 0.783 |
| Esophagus | −0.243 | −0.144 | 0.564 | 0.682 | 0.863 |
| Stomach | −0.778 | −0.215 | −0.266 | 0.455 | 0.930 |
| Intestine | 0.630 | 0.682 | 0.194 | −0.115 | 0.913 |
| Rectum | 0.130 | 0.991 | −0.019 | −0.012 | 0.999 |
| Liver | −0.755 | −0.156 | 0.240 | −0.022 | 0.653 |
| Pancreas | 0.841 | 0.075 | −0.129 | 0.182 | 0.762 |
| Larynx | −0.194 | −0.019 | 0.835 | 0.085 | 0.742 |
| Lung | 0.187 | 0.586 | −0.033 | 0.273 | 0.454 |
| Skin | 0.511 | 0.083 | 0.166 | −0.123 | 0.311 |
| Thyroid | 0.021 | 0.110 | 0.065 | 0.351 | 0.140 |
| Leukemia | 0.547 | 0.221 | −0.107 | −0.198 | 0.399 |
| Prostate | 0.722 | 0.142 | 0.070 | 0.125 | 0.562 |
| Sum of squares | 3.630 | 1.982 | 1.877 | 1.022 | 8.511 |
| Contribution rate | 27.9% | 15.2% | 14.4% | 7.9% | 65.5% |

**Table 11. Factor scores of 24 countries for males by 13 sites of cancer.**

| Country | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| Union of South Africa | 1.030 | −0.774 | 0.270 | 0.433 |
| Canada | 0.995 | −0.024 | −0.133 | −0.415 |
| Chile | −1.082 | −1.710 | −1.017 | 2.051 |
| United States, white | 1.423 | −0.477 | 0.223 | −0.647 |
| United States, nonwhite | 1.291 | −1.004 | 0.343 | 0.750 |
| Israel | 0.187 | −1.424 | −0.368 | −1.290 |
| Japan | −2.883 | −0.510 | −0.535 | 0.086 |
| Germany Federal Republic | −1.060 | 0.625 | −0.689 | −0.497 |
| Austria | −0.529 | 1.059 | −0.374 | 0.660 |
| Belgium | −0.856 | 1.140 | 0.190 | −1.021 |
| Denmark | −0.162 | 2.275 | −1.153 | 0.224 |
| Finland | −0.534 | −0.677 | −0.471 | 1.860 |
| France | −0.444 | 0.104 | 3.270 | 0.843 |
| Ireland | 0.048 | 0.652 | 0.656 | −0.407 |
| Italy | −1.144 | −0.866 | 1.182 | −1.494 |
| Norway | 0.416 | −0.979 | −0.951 | −0.147 |
| Netherlands | −0.220 | 0.012 | −0.691 | −0.515 |
| Portugal | −0.249 | −1.178 | 0.586 | −1.073 |
| England and Wales | 0.066 | 1.601 | −0.245 | 0.167 |
| Scotland | 0.681 | 1.336 | 0.242 | 0.701 |
| Northern Ireland | 0.249 | 0.762 | −0.114 | −0.208 |
| Sweden | 0.591 | −0.309 | −1.112 | −0.355 |
| Switzerland | 0.197 | 0.380 | 1.344 | 1.580 |
| Australia | 1.028 | −0.330 | −0.046 | −0.934 |
| New Zealand | 0.964 | 0.319 | −0.405 | −0.352 |

Table 12. Factor loadings of 14 sites of cancer for females.

| Site | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Communality |
|---|---|---|---|---|---|
| Oral | 0.390 | 0.268 | 0.809 | 0.053 | 0.881 |
| Esophagus | −0.199 | −0.251 | 0.676 | 0.327 | 0.667 |
| Stomach | −0.640 | −0.568 | 0.104 | 0.271 | 0.816 |
| Intestine | 0.371 | 0.887 | 0.197 | −0.186 | 0.998 |
| Rectum | 0.003 | 0.702 | 0.013 | 0.049 | 0.495 |
| Liver | −0.457 | −0.246 | −0.545 | 0.131 | 0.583 |
| Pancreas | 0.295 | 0.491 | 0.304 | 0.406 | 0.586 |
| Larynx | −0.100 | 0.073 | 0.714 | −0.100 | 0.534 |
| Lung | −0.110 | 0.514 | 0.399 | 0.447 | 0.635 |
| Skin | 0.567 | 0.246 | −0.029 | 0.096 | 0.393 |
| Thyroid | 0.069 | −0.001 | −0.061 | 0.611 | 0.382 |
| Leukemia | 0.659 | 0.256 | −0.565 | 0.362 | 0.950 |
| Breast | 0.482 | 0.805 | −0.087 | 0.188 | 0.924 |
| Uterus | −0.534 | 0.010 | −0.075 | −0.033 | 0.292 |
| Sum of squares | 2.335 | 3.082 | 2.555 | 1.163 | 9.136 |
| Contribution rate | 16.7% | 22.0% | 18.3% | 8.3% | 65.3% |

Table 13. Factor scores of 24 countries for females by 14 sites of cancer.

| Country | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| Union of South Africa | 0.562 | 0.196 | −0.487 | 1.168 |
| Canada | 0.246 | 1.622 | −0.339 | −1.006 |
| Chile | −2.065 | −1.023 | 0.666 | 1.084 |
| United States, white | 1.082 | 0.582 | −0.747 | −0.248 |
| United States, nonwhite | −0.417 | 0.613 | 0.675 | −0.614 |
| Israel | 0.736 | −0.558 | −1.329 | 2.241 |
| Japan | −1.995 | −2.020 | −0.126 | −1.116 |
| German Federal Republic | −1.492 | −0.513 | −0.349 | −0.153 |
| Austria | −1.136 | 0.354 | −0.577 | 1.112 |
| Belgium | −0.443 | 0.953 | −1.169 | −0.794 |
| Denmark | 0.262 | 0.805 | −1.416 | 1.422 |
| Finland | 0.638 | −1.756 | 1.199 | 0.912 |
| France | 0.329 | −0.252 | −1.120 | −0.977 |
| Ireland | 0.690 | −0.095 | 2.107 | 0.526 |
| Italy | 0.444 | −0.501 | −0.541 | −0.976 |
| Norway | 1.553 | −1.144 | −0.724 | 0.298 |
| Netherlands | −0.397 | 0.640 | −0.812 | 0.067 |
| Portugal | −0.699 | −1.289 | 0.403 | −1.393 |
| England and Wales | −0.663 | 0.687 | 0.683 | 0.256 |
| Scotland | −0.642 | 2.206 | 1.996 | −0.186 |
| Northern Ireland | 0.482 | 0.393 | 1.858 | −0.117 |
| Sweden | 0.737 | −0.570 | 0.266 | 0.320 |
| Switzerland | 0.299 | −0.063 | −0.784 | 0.377 |
| Australia | 0.735 | 0.161 | 0.328 | −1.556 |
| New Zealand | 1.154 | 0.573 | 0.340 | −0.643 |

***Second Factor.*** In both sexes, rectum, intestine and lung cancers have commonly high loadings on this factor, and breast cancer has also high loadings in females. It is presumed, therefore, that the second factor functions in males as well as in females in a similar way in that considerably high positive loading are found for intestine, rectum and lung cancers. It should be noted, however, that some discrepancies are found in the locations of factor scores shown in the vertical lines of Figure 5 and 7. In males, factor scores of Denmark are highest, followed by England and Wales, and Scotland, while in females Scotland is highest, followed by Canada and Denmark. Note that Japan is plotted lowest on this factor.

***Third Factor.*** Oral, esophagus, and larynx cancers have high factor loadings on this factor in both sexes. On the contrary, distributions of factor scores on both sexes are dissimilar, since France and Switzerland, where high positive factor scores are found in males, have negative factor scores in females.

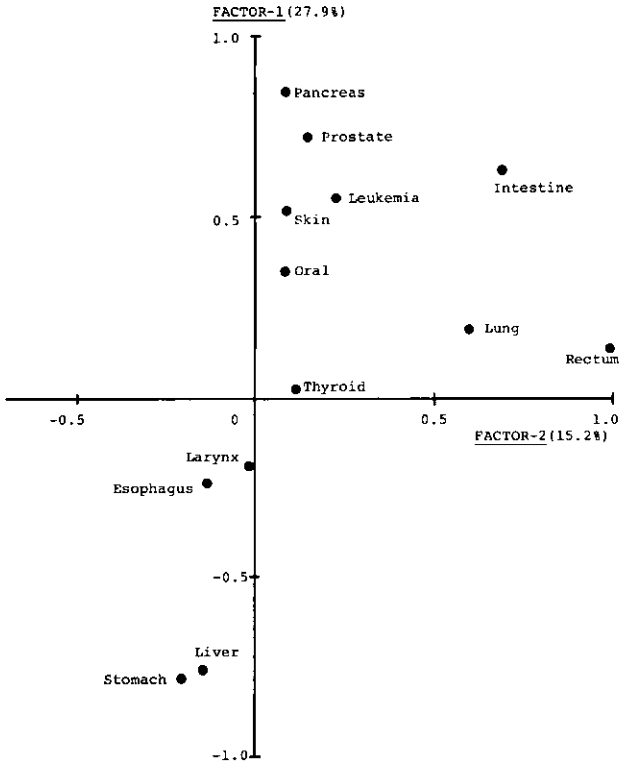


FIGURE 1. Configuration of 13 sites of cancer by factor loadings for males (factor 1 and factor 2).

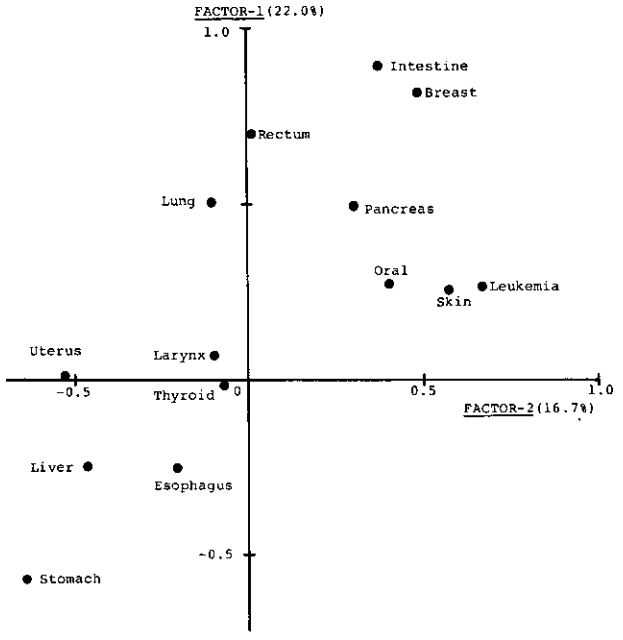


FIGURE 3. Configuration of 14 sites of cancer by factor loadings for females (factor 1 and factor 2).

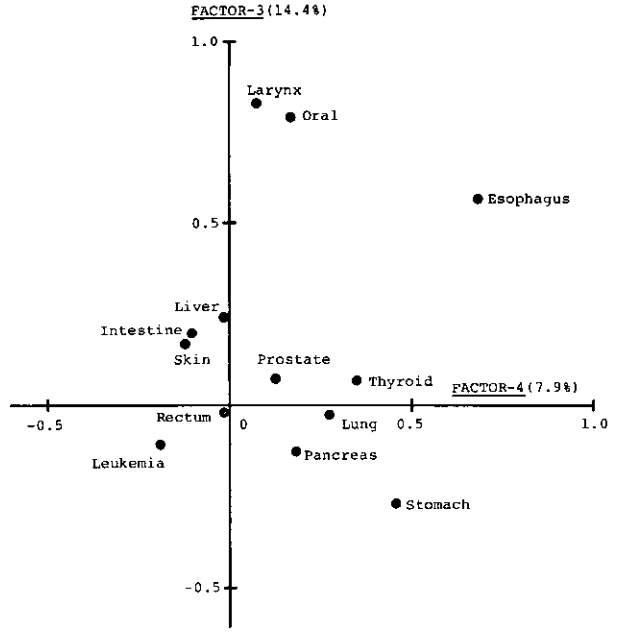


FIGURE 2. Configuration of 13 sites of cancer by factor loadings for males (factor 3 and factor 4).

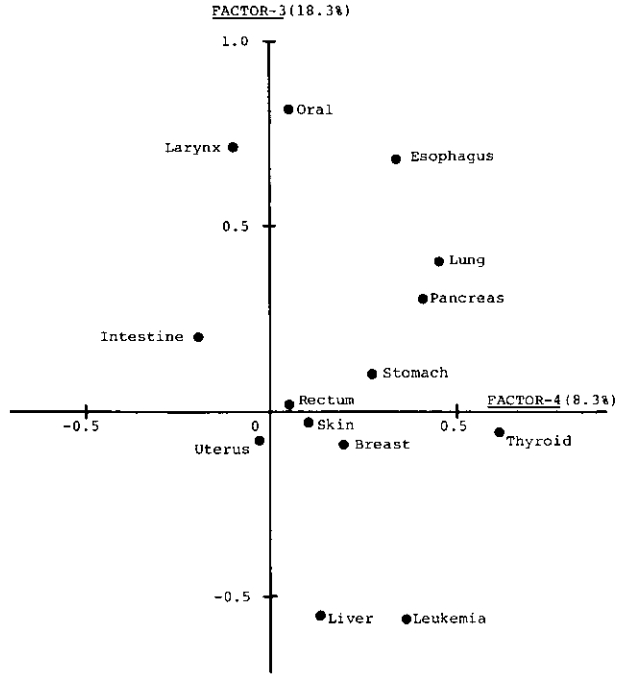


FIGURE 4. Configuration of 14 sites of cancer by factor loadings for females (factor 3 and factor 4).

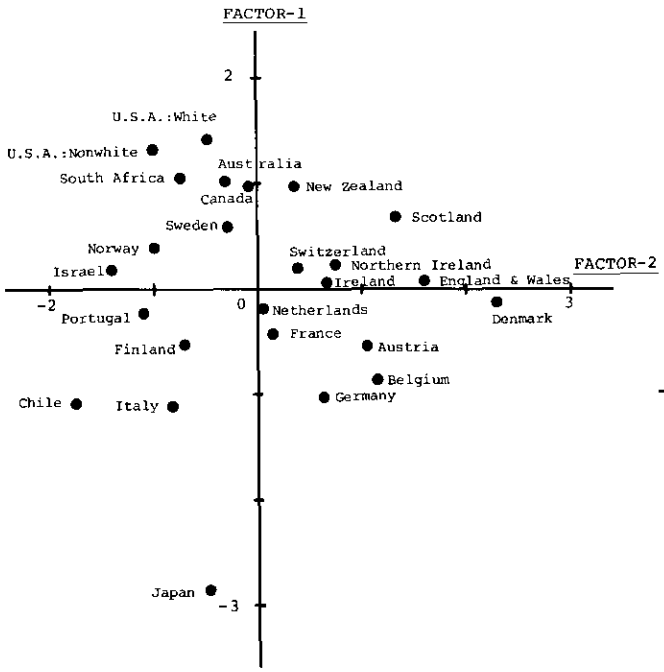


FIGURE 5. Configuration of 24 countries by factor scores for males by 13 sites of cancer (factor 1 and factor 2).

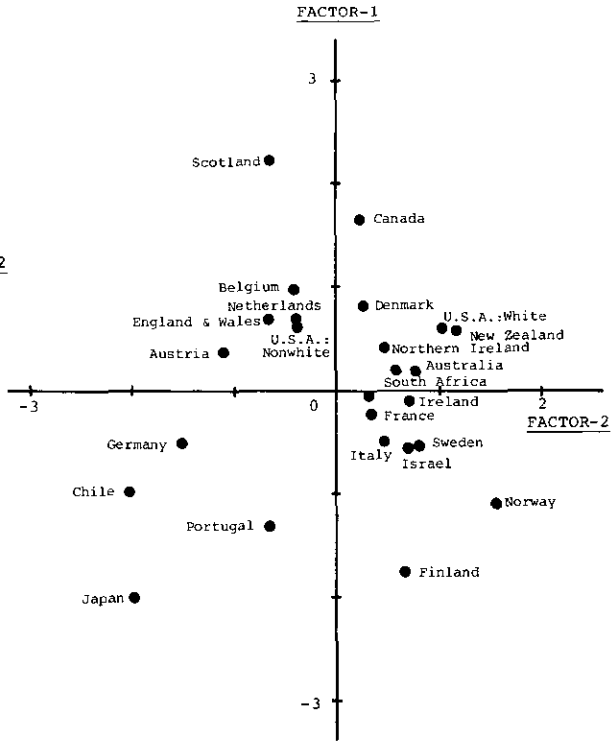


FIGURE 7. Configuration of 24 countries by factor scores for females by 14 sites of cancer (factor 1 and factor 2).

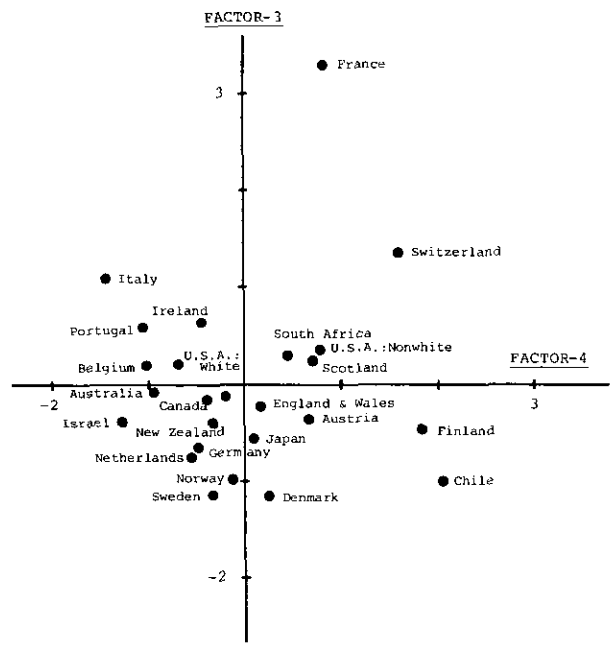


FIGURE 6. Configuration of 24 countries by factor scores for males by 13 sites of cancer (factor 3 and factor 4).

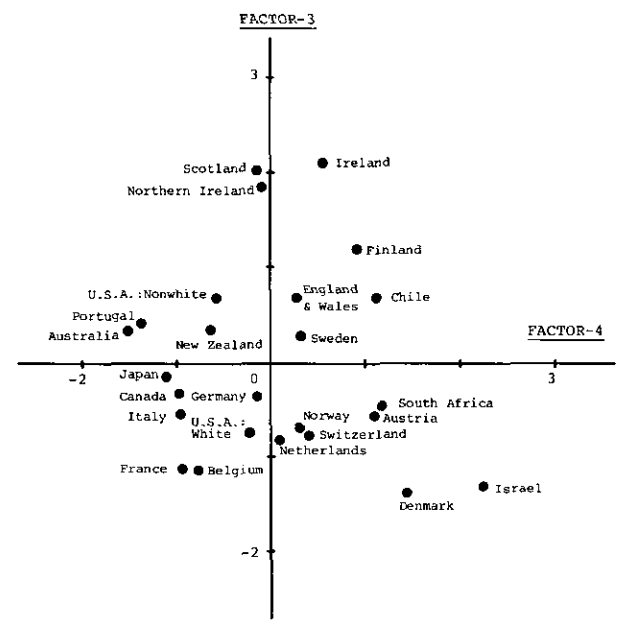


FIGURE 8. Configuration of 24 countries by factor scores for females by 14 sites of cancer (factor 3 and factor 4).

***Fourth Factor.*** High loadings were found in esophagus and stomach cancers in males, while relatively high loadings were found in thyroid and lung cancers, showing quite dissimilar patterns with regard to sex. In view of its relatively small contribution ratios (7.9% in males and 8.3% in females) the fourth factor does not seem to bear any substantial meanings behind it.

In order to substantiate the meaning of the four factors, we calculated the correlation coefficients between these factor scores and the scores on the 24 variables shown in Table 3. The results for males are shown in Table 14 and those for females in Table 15.

The first factor correlates highly with intakes of sugar, meat, spices, and animal products in the positive direction and with population density in the negative direction in case of males, while it correlates highly and positively milk, sugar, meat, and energy intake and negatively with population density in the case of females. It is noted for both sexes that the first factor correlates positively with animal products and negatively with vegetable products. This tendency is more obvious in males rather than in females. The second factor has, on the whole, a similar pattern of correlations as the first factor but oil fats and animal oil have higher correlations to this factor than the first factor. It is interesting to note that the second factor has a strong negative correlation with population density in both sexes, suggesting that cancers of intestine, rectum, and lung are liable to occur more frequently in areas where population is not so dense. For the third factor, large discrepancies are found between males and females as they are seen between the distributions of factor scores. In males, the third factor has strong correlations with intake of alcohol, vegetables, and vegetable products, while it has no significant correlations with these variables especially in females. For the fourth factor, we may only note that rainfall and vegetable have high negative correlations with this factor only in females. Next, we present in Table 16 the results of factor analysis based on the mortality rates of 27 sites of males and females. From Table 16 we see that the first and second factors correspond to those of Table 10 and Table 12 in terms of the sites with high factor loadings. It is interesting to note that the third factor which is shown in Table 16 corresponds to the third factor of males shown in Table 10 and the fourth factor corresponds to the third factor in Table 12. This result implies that the geographical distributions of oral, esophagus, and larynx cancers are different with respect to each sex, thus forming different clusters. Scrutinizing the result of Table 16, we note that the correspondence mentioned above will not be so clear in the strict sense, since pancreas cancer in both sexes have high positive factor loadings, and larynx cancer of females have relatively lower factor loadings as compared with that of Table 10.

**Table 14. Correlation coefficients between each item and each factor for males.**

| Item | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| Primary industry | −0.309 | −0.469 | 0.039 | 0.177 |
| Population density | −0.576 | 0.352 | −0.011 | −0.254 |
| Mean temperature | 0.044 | −0.490 | 0.237 | −0.236 |
| Rainfall | 0.037 | 0.033 | 0.037 | −0.040 |
| Alcohol | −0.211 | 0.225 | 0.685 | −0.011 |
| Energy | 0.420 | 0.638 | 0.238 | −0.024 |
| Vegetable products | −0.470 | −0.140 | 0.411 | −0.107 |
| Animal products | 0.539 | 0.507 | −0.056 | 0.035 |
| Cereals | −0.468 | −0.387 | 0.157 | 0.067 |
| Roots and tubers | −0.325 | 0.510 | −0.042 | 0.016 |
| Sugar | 0.737 | 0.312 | −0.176 | 0.145 |
| Pulses | −0.236 | −0.522 | 0.164 | −0.182 |
| Tree nuts | −0.306 | −0.238 | 0.056 | −0.206 |
| Vegetables | −0.186 | −0.301 | 0.503 | −0.250 |
| Fruits | 0.120 | −0.077 | 0.144 | −0.347 |
| Meat | 0.635 | 0.448 | 0.075 | −0.086 |
| Eggs | 0.356 | 0.303 | −0.028 | −0.299 |
| Fish | −0.473 | −0.154 | −0.232 | −0.083 |
| Milk | 0.497 | 0.309 | −0.170 | 0.194 |
| Oil fats | 0.202 | 0.586 | −0.046 | −0.154 |
| Vegetable oils | −0.056 | 0.086 | −0.018 | −0.288 |
| Animal oils | 0.288 | 0.605 | −0.038 | 0.087 |
| Spices | 0.583 | 0.254 | −0.102 | 0.168 |
| Stimulants | 0.267 | 0.338 | 0.032 | 0.136 |

**Table 15. Correlation coefficients between each item and each factor for females.**

| Item | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| Primary industry | −0.043 | −0.659 | 0.026 | 0.179 |
| Population density | −0.466 | 0.019 | −0.288 | −0.113 |
| Mean temperature | −0.163 | −0.115 | −0.070 | −0.186 |
| Rainfall | −0.074 | 0.053 | 0.108 | −0.615 |
| Alcohol | −0.202 | 0.078 | −0.191 | −0.397 |
| Energy | 0.407 | 0.518 | −0.004 | −0.069 |
| Vegetable products | −0.202 | −0.313 | −0.286 | 0.107 |
| Animal products | 0.382 | 0.530 | 0.145 | −0.111 |
| Cereals | −0.280 | −0.467 | 0.047 | 0.097 |
| Roots and tubers | −0.122 | −0.062 | 0.148 | −0.034 |
| Sugar | 0.468 | 0.551 | 0.100 | 0.225 |
| Pulses | −0.334 | −0.270 | 0.064 | −0.214 |
| Tree nuts | −0.271 | −0.204 | −0.171 | −0.253 |
| Vegetables | −0.191 | −0.004 | −0.125 | −0.465 |
| Fruits | 0.110 | 0.084 | −0.284 | −0.048 |
| Meat | 0.323 | 0.652 | 0.073 | −0.258 |
| Eggs | 0.233 | 0.493 | −0.079 | 0.036 |
| Fish | −0.175 | −0.554 | −0.052 | −0.228 |
| Milk | 0.552 | 0.257 | 0.166 | 0.175 |
| Oil fats | 0.231 | 0.435 | −0.236 | 0.039 |
| Vegetable oils | 0.125 | 0.020 | −0.507 | 0.179 |
| Animal oils | 0.153 | 0.489 | 0.196 | −0.122 |
| Spices | 0.263 | 0.459 | −0.139 | 0.151 |
| Stimulants | 0.246 | 0.327 | −0.251 | 0.165 |

**Table 16. Factor loadings of 13 sites in males and 14 sites in females of cancer.**

| Site | I | II | III | IV | Communality |
|---|---|---|---|---|---|
| Male | | | | | |
| Oral | 0.355 | 0.044 | 0.903 | 0.227 | 0.995 |
| Esophagus | −0.292 | −0.133 | 0.600 | −0.084 | 0.543 |
| Stomach | −0.859 | −0.183 | −0.042 | −0.243 | 0.832 |
| Intestine | 0.596 | 0.680 | 0.169 | 0.315 | 0.944 |
| Rectum | 0.226 | 0.909 | −0.020 | −0.008 | 0.878 |
| Liver | −0.431 | −0.126 | 0.197 | −0.771 | 0.834 |
| Pancreas | 0.439 | 0.122 | −0.102 | 0.715 | 0.732 |
| Larynx | 0.089 | −0.060 | 0.705 | −0.374 | 0.558 |
| Lung | 0.005 | 0.683 | 0.041 | 0.235 | 0.415 |
| Skin | 0.591 | 0.010 | 0.168 | 0.200 | 0.861 |
| Thyroid | 0.029 | 0.057 | 0.110 | −0.106 | 0.024 |
| Leukemia | 0.781 | 0.134 | −0.400 | −0.033 | 0.446 |
| Prostate | 0.628 | 0.107 | −0.025 | 0.242 | 0.734 |
| Female | | | | | |
| Oral | 0.070 | 0.248 | 0.111 | 0.815 | 0.637 |
| Esophagus | −0.625 | 0.003 | 0.078 | 0.519 | 0.884 |
| Stomach | −0.875 | −0.155 | −0.074 | −0.176 | 0.878 |
| Intestine | 0.606 | 0.605 | 0.094 | 0.364 | 0.872 |
| Rectum | 0.193 | 0.930 | −0.162 | −0.004 | 0.672 |
| Liver | −0.270 | −0.059 | −0.106 | −0.744 | 0.701 |
| Pancreas | 0.321 | 0.197 | −0.129 | 0.721 | 0.623 |
| Larynx | −0.311 | 0.251 | 0.513 | 0.372 | 0.378 |
| Lung | −0.080 | 0.538 | 0.004 | 0.432 | 0.457 |
| Skin | 0.617 | 0.099 | 0.164 | 0.167 | 0.707 |
| Thyroid | −0.006 | 0.106 | 0.019 | 0.105 | 0.120 |
| Leukemia | 0.729 | 0.044 | −0.450 | −0.080 | 0.813 |
| Breast | 0.700 | 0.528 | −0.170 | 0.278 | 0.874 |
| Uterus | −0.296 | −0.046 | 0.029 | −0.104 | 0.101 |

Table 17. Results of stepwise multiple regression analysis for males.

| Site | Order | Item | $\beta$ | Correlative coefficient | Multiple correlation coefficient |
|---|---|---|---|---|---|
| Oral | 1 | Alcohol | 0.810 | ( 0.456) | 0.456 |
| | 2 | Population density | −0.263 | (−0.300) | 0.596 |
| | 3 | Fruits | −0.207 | (−0.013) | 0.648 |
| | 4 | Sugar | 0.688 | ( 0.150) | 0.702 |
| | 5 | Cereals | 0.489 | ( 0.065) | 0.765 |
| | | | | 0.476 | 0.586 |
| Esophagus | 1 | Alcohol | 0.624 | ( 0.398) | 0.398 |
| | 2 | Fruits | −0.348 | (−0.086) | 0.490 |
| | 3 | Tree nuts | 0.197 | ( 0.096) | 0.523 |
| | 4 | Population density | −0.218 | (−0.042) | 0.555 |
| | 5 | Meat | −0.173 | (−0.160) | 0.579 |
| | | | | 0.158 | 0.334 |
| Stomach | 1 | Meat | −0.706 | (−0.659) | 0.659 |
| | 2 | Rainfall | 0.465 | (−0.041) | 0.698 |
| | 3 | Tree nuts | −0.306 | ( 0.149) | 0.744 |
| | 4 | Temperature | −0.205 | (−0.085) | 0.753 |
| | 5 | Energy | −0.297 | (−0.585) | 0.767 |
| | | | | 0.479 | 0.587 |
| Intestine | 1 | Meat | 1.092 | ( 0.731) | 0.731 |
| | 2 | Rainfall | −0.292 | ( 0.120) | 0.752 |
| | 3 | Primary industry | −0.452 | (−0.605) | 0.777 |
| | 4 | Cereals | 0.680 | (−0.501) | 0.811 |
| | 5 | Temperature | −0.321 | (−0.243) | 0.854 |
| | | | | 0.657 | 0.728 |
| Rectum | 1 | Energy | 0.857 | ( 0.707) | 0.707 |
| | 2 | Population density | 0.130 | ( 0.313) | 0.765 |
| | 3 | Temperature | −0.643 | (−0.509) | 0.805 |
| | 4 | Milk | −0.977 | ( 0.403) | 0.844 |
| | 5 | Sugar | 0.521 | ( 0.428) | 0.896 |
| | | | | 0.750 | 0.802 |
| Liver | 1 | Sugar | −0.809 | (−0.759) | 0.759 |
| | 2 | Population density | 0.256 | ( 0.446) | 0.799 |
| | 3 | Spices | 0.203 | (−0.279) | 0.830 |
| | 4 | Tree nuts | 0.227 | ( 0.524) | 0.841 |
| | 5 | Fish | −0.191 | ( 0.319) | 0.854 |
| | | | | 0.659 | 0.730 |
| Pancreas | 1 | Sugar | 0.502 | ( 0.664) | 0.664 |
| | 2 | Population density | −0.498 | (−0.634) | 0.814 |
| | 3 | Milk | −0.241 | ( 0.444) | 0.828 |
| | 4 | Meat | 0.309 | ( 0.560) | 0.839 |
| | 5 | Rainfall | −0.161 | (−0.047) | 0.850 |
| | | | | 0.650 | 0.722 |
| Larynx | 1 | Alcohol | 0.878 | ( 0.760) | 0.760 |
| | 2 | Fruits | −0.368 | ( 0.112) | 0.797 |
| | 3 | Vegetables | 0.181 | ( 0.474) | 0.823 |
| | 4 | Primary industry | 0.476 | ( 0.271) | 0.845 |
| | 5 | Eggs | 0.402 | (−0.165) | 0.878 |
| | | | | 0.710 | 0.771 |

Table 17 (Cont'd)

| Site | Order | Item | $\beta$ | Correlative coefficient | Multiple correlation coefficient |
|---|---|---|---|---|---|
| Lung | 1 | Energy | 0.224 | ( 0.465) | 0.465 |
| | 2 | Vegetable | −0.315 | (−0.363) | 0.535 |
| | 3 | Fish | −0.425 | (−0.416) | 0.606 |
| | 4 | Population density | 0.308 | ( 0.209) | 0.674 |
| | 5 | Fruits | −0.143 | (−0.078) | 0.686 |
| | | | | 0.331 | 0.471 |
| Skin | 1 | Fish | −0.079 | (−0.404) | 0.404 |
| | 2 | Population density | −0.195 | (−0.386) | 0.515 |
| | 3 | Spices | 0.192 | ( 0.400) | 0.573 |
| | 4 | Cereals | 0.829 | ( 0.106) | 0.649 |
| | 5 | Meat | 0.826 | ( 0.381) | 0.766 |
| | | | | 0.478 | 0.587 |
| Thyroid | 1 | Fruits | 0.712 | ( 0.479) | 0.479 |
| | 2 | Vegetable | −0.163 | (−0.240) | 0.636 |
| | 3 | Cereals | 0.595 | ( 0.039) | 0.711 |
| | 4 | Spices | 0.462 | ( 0.389) | 0.769 |
| | 5 | Temperature | −0.409 | (−0.219) | 0.821 |
| | | | | 0.587 | 0.693 |
| Leukemia | 1 | Sugar | −0.078 | ( 0.570) | 0.570 |
| | 2 | Spices | 0.590 | ( 0.555) | 0.633 |
| | 3 | Rainfall | −0.462 | (−0.183) | 0.732 |
| | 4 | Cereals | −0.239 | (−0.493) | 0.766 |
| | 5 | Milk | 0.211 | ( 0.543) | 0.775 |
| | | | | 0.495 | 0.600 |
| Prostate | 1 | Spices | 0.622 | ( 0.680) | 0.680 |
| | 2 | Tree nuts | −0.394 | (−0.437) | 0.812 |
| | 3 | Temperature | 0.390 | (−0.088) | 0.835 |
| | 4 | Cereals | −0.560 | (−0.540) | 0.880 |
| | 5 | Eggs | −0.288 | ( 0.207) | 0.897 |
| | | | | 0.753 | 0.805 |

## Stepwise Regression

Next, we show the results of stepwise regression for males in Table 17 and of females in Table 18. In each case, the scores are calculated for each of the five steps. The item number indicates the selected order of the item and the table gives the multiple correlation coefficient when the items, the number of which is equal to its order, are employed in the regression equation. Let us illustrate this for the case of oral cancer in males. In the first step, alcohol is selected. The population density is added in the second step, producing the multiple correlation coefficient of 0.596. Continuing the steps in the same way, we come to the final step where five items as alcohol, population density, fruits, sugar and cereals are in the equation. We list the partial regression coefficients, under the heading beta coefficient, for the five items selected in the final stage in each cancer site. Suppose that each variable is standardized with zero mean and unit variance, then the mortality of oral cancer ($\hat{y}$) for each country is predicted by $\hat{y} = 0.81 \times \text{alcohol} - 0.263 \times \text{population density} - 0.207 \times \text{fruit} + 0.688 \times \text{sugar} + 0.488 \times \text{cereal}$, which yields a multiple correlation coefficient of 0.765.

The squared multiple correlation coefficient (SMC) and the squared multiple correlation coefficient (SMCR) adjusted with regard to the degree of freedom are also shown in Tables 17 and 18. The number of statistically significant SMC at the 1% level (SMC $\geqq$ 0.522) are 11 and 9 in males and females, respectively. The largest multiple correlation coefficients were observed in prostate cancer in

**Table 18. Results of stepwise multiple regression analysis for females.**

| Site | Order | Item | $\beta$ | Correlative coefficient | Multiple correlation coeffecient |
|---|---|---|---|---|---|
| Oral | 1 | Milk | 0.433 | ( 0.513) | 0.513 |
| | 2 | Population density | −0.374 | (−0.394) | 0.579 |
| | 3 | Fruits | −0.265 | (−0.236) | 0.629 |
| | 4 | Spices | −0.200 | ( 0.153) | 0.639 |
| | 5 | Primary industry | −0.201 | (−0.156) | 0.659 |
| | | | | 0.285 | 0.436 |
| Esophagus | 1 | Spices | −0.570 | (−0.402) | 0.402 |
| | 2 | Alcohol | −0.364 | (−0.362) | 0.592 |
| | 3 | Fish | −0.288 | (−0.044) | 0.647 |
| | 4 | Fruits | −0.262 | (−0.372) | 0.688 |
| | 5 | Population density | −0.200 | (−0.185) | 0.713 |
| | | | | 0.379 | 0.508 |
| Stomach | 1 | Meat | −0.748 | (−0.657) | 0.657 |
| | 2 | Temperature | −0.408 | (−0.116) | 0.682 |
| | 3 | Energy | −0.334 | (−0.561) | 0.701 |
| | 4 | Fish | −0.376 | ( 0.292) | 0.712 |
| | 5 | Rainfall | 0.331 | (−0.164) | 0.750 |
| | | | | 0.447 | 0.561 |
| Intestine | 1 | Meat | 0.931 | ( 0.749) | 0.749 |
| | 2 | Rainfall | −0.298 | ( 0.133) | 0.768 |
| | 3 | Fruits | −0.155 | (−0.059) | 0.787 |
| | 4 | Primary industry | −0.319 | (−0.564) | 0.801 |
| | 5 | Cereals | 0.347 | (−0.498) | 0.824 |
| | | | | 0.594 | 0.678 |
| Rectum | 1 | Energy | 0.502 | ( 0.664) | 0.664 |
| | 2 | Population density | 0.243 | ( 0.329) | 0.732 |
| | 3 | Fruits | −0.215 | (−0.084) | 0.773 |
| | 4 | Primary industry | −0.248 | (−0.576) | 0.796 |
| | 5 | Temperature | −0.192 | (−0.459) | 0.814 |
| | | | | 0.573 | 0.662 |
| Liver | 1 | Sugar | −0.819 | (−0.580) | 0.580 |
| | 2 | Population density | 0.449 | ( 0.523) | 0.693 |
| | 3 | Fish | −0.332 | ( 0.008) | 0.773 |
| | 4 | Spices | 0.360 | (−0.229) | 0.798 |
| | 5 | Rainfall | −0.236 | (−0.223) | 0.823 |
| | | | | 0.592 | 0.671 |
| Pancreas | 1 | Sugar | 0.254 | ( 0.633) | 0.633 |
| | 2 | Population density | −0.436 | (−0.551) | 0.746 |
| | 3 | Rainfall | −0.394 | (−0.228) | 0.787 |
| | 4 | Eggs | 0.280 | ( 0.367) | 0.820 |
| | 5 | Spices | 0.250 | ( 0.449) | 0.841 |
| | | | | 0.629 | 0.707 |
| Larynx | 1 | Fruits | −0.527 | (−0.346) | 0.346 |
| | 2 | Spices | −0.505 | (−0.301) | 0.451 |
| | 3 | Fish | −0.389 | (−0.093) | 0.524 |
| | 4 | Tree nuts | 0.334 | (−0.027) | 0.559 |
| | 5 | Energy | 0.285 | (−0.008) | 0.600 |
| | | | | 0.190 | 0.359 |

**Table 18 (Cont'd).**

| Site | Order | Item | $\beta$ | Correlative coefficient | Multiple correlation coefficient |
|---|---|---|---|---|---|
| Lung | 1 | Fish | −0.248 | (−0.378) | 0.378 |
| | 2 | Alcohol | −0.222 | (−0.170) | 0.453 |
| | 3 | Rainfall | −0.469 | (−0.321) | 0.512 |
| | 4 | Primary industry | −0.753 | (−0.299) | 0.614 |
| | 5 | Cereals | 0.506 | ( 0.008) | 0.715 |
| | | | | 0.382 | 0.511 |
| Skin | 1 | Fish | 0.075 | (−0.453) | 0.453 |
| | 2 | Spices | 0.289 | ( 0.386) | 0.534 |
| | 3 | Tree nuts | −0.474 | (−0.436) | 0.609 |
| | 4 | Cereals | 0.826 | ( 0.019) | 0.683 |
| | 5 | Meat | 0.785 | ( 0.405) | 0.786 |
| | | | | 0.517 | 0.618 |
| Thyroid | 1 | Fruits | 0.707 | ( 0.439) | 0.439 |
| | 2 | Vegetables | −0.515 | (−0.307) | 0.647 |
| | 3 | Cereals | 0.587 | ( 0.106) | 0.762 |
| | 4 | Fish | −0.305 | (−0.292) | 0.799 |
| | 5 | Temperature | −0.293 | (−0.216) | 0.827 |
| | | | | 0.600 | 0.684 |
| Leukemia | 1 | Milk | 0.286 | ( 0.482) | 0.482 |
| | 2 | Rainfall | −0.437 | (−0.260) | 0.576 |
| | 3 | Spices | 0.393 | (0.405) | 0.676 |
| | 4 | Fruits | 0.365 | (0.314) | 0.727 |
| | 5 | Alcohol | −0.217 | (−0.175) | 0.751 |
| | | | | 0.449 | 0.564 |
| Breast | 1 | Sugar | 0.585 | ( 0.764) | 0.764 |
| | 2 | Population density | 0.256 | (−0.006) | 0.792 |
| | 3 | Fish | −0.199 | (−0.556) | 0.821 |
| | 4 | Spices | 0.275 | ( 0.538) | 0.834 |
| | 5 | Rainfall | −0.180 | (−0.085) | 0.848 |
| | | | | 0.646 | 0.720 |
| Uterus | 1 | Temperature | 0.523 | ( 0.349) | 0.349 |
| | 2 | Spices | 0.427 | ( 0.177) | 0.415 |
| | 3 | Sugar | −0.650 | (−0.125) | 0.462 |
| | 4 | Cereals | −0.558 | ( 0.043) | 0.508 |
| | 5 | Fruits | −0.333 | (−0.082) | 0.585 |
| | | | | 0.169 | 0.342 |

males ($r$ = 0.897) and breast cancer in females ($r$ = 0.843). Cancers of the intestine, rectum, liver, pancreas, and thyroid also have larger multiple correlations, as compared to the other sites of cancer in both sexes. Next, we shall describe some interesting points for each site of cancer concerning the result of stepwise regression procedure as shown in Table 17 and 18. First, oral, esophagus, and larynx cancers in males, which are found to form a cluster with regard to their geographical distribution, have high positive regression coefficients on alcohol and negative coefficients on fruit, but in case of females only fruit has a negative regression coefficient for each of the three cases. Secondly, for the regression coefficients of stomach cancer, meat has large negative values and rainfall has positive values in both sexes, but they are opposite in sign for intestine cancer although the magnitude of correlation coefficient between rainfall and intestine cancer is negligible. Thirdly, it is noteworthy that signs of the regression coefficients of sugar and population density are opposite with respect to liver and pancreas cancers in both sexes. Finally, it is interesting to note that fruit has large value of regression coefficients on thyroid cancer in both sexes and that spices have high value with respect to prostate cancer.

# Discussion

## Correlation Coefficients of the Geographical Distributions

The trend of site-specific mortality rates of cancer is shown in Table 4. In most cases of the sites of cancer, the magnitude of mortalities is relatively stable during these three periods. However, as we have already mentioned, the mortality rate of lung cancer has been increasing and that of stomach cancer has been decreasing in both sexes over the period from 1950 through 1966. From Table 4, we can easily calculate the sum of mortality rates of 13 male cancers in three periods, which amount to 114.73, 131.83, and 126.6 per 100,000 population and those of 14 female cancers to 92.87, 89.50, and 86.37. Thus a consistent decreasing trend is observed in female cancers from the first to third periods, while the male cancers have a peak in the second period. As to the sex ratios shown in the third column of Table 4, we note that high values are found in oral, esophagus, larynx, and lung cancers, in which the cigarette smoking is suspected as one of the etiological factors. Next, we shall consider the result of the geographical distributions in the different periods (Tables 5 and 6). Consistently high correlation coefficients were obtained for most of the cancer sites. The result would surely tend to support Burkitt's assumption that environmental factors implicated through the geographical distribution might function as the predominant factors in cancer etiology together with the hypothesis that race differences might be related to cancer etiology in some way. As for the interpretation of the magnitude of correlation coefficients, we should keep in mind to recognize the fact that correlation coefficients are easily biased by the way sampling is done from the population. It is well known that correlations between two cancer sites obtained from the countries over the world sometimes differ drastically from those obtained from the samples of Japanese prefectures. For example, we note from Tables 8 and 9 that a high negative correlation is observed between the international distribution of stomach cancer and intestine cancer ($r = -0.729$ for males, $r = -0.733$ for females).

However, according to the result based on the Japanese data (*9*), they are positively correlated in both sexes ($r = 0.436$ for males, $r = 0.425$ for females). In view of this disagreement we should take some care about nature of the sample used for calculating the correlation coefficient when we interprete the magnitude of correlation coefficients and the factor loadings for each variable.

## Factor Analysis

Now, we shall give some comments on the results of factor analysis, which are shown in Tables 10 through 16 and Fig. 1 through 8. Since the results of factor analysis indicate that the various cancer sites formed clusters with respect to their geographical distributions over the periods examined, we may postulate the existence of some common causes to develop cancer in respective sites. It should be noted, however, that similarities in the geographical distributions of two diseases may sometimes reflect only statistical bias in sampling. Hence, it is better to develop an etiological hypothesis about alternative diseases within a cluster only when we have some established hypothesis about the etiology of a certain disease in the cluster. In view of this, we shall consider what clues each factor provides to develop an etiological hypothesis, using the result of factor analysis as well as that of stepwise regression method. It is fairly obvious that the correlations between the mortality rates of some cancers (prostate, pancreas, skin, and intestine) and the first factor are very strong in U.S. males and very low in Japan. One of these reasons may be attributable to the difference observed in the amount of intake of sugar, meat, milk, and animal products which have high factor loadings on this factor. This fact conforms to the finding reported by Lea (*15*), who also analyzed interrelationships between the mortality rates of some cancer sites and food intake. Hence, in view of the findings, we might say that the first factor may be related to the Western style food habits. Unfortunately, we cannot give any satisfactory explanations for the close relationship between foods habits and the high mortality rate in skin and prostate cancers, and leukemia which have high factor loadings on the first factor of males. In connection between prostate cancer and leukemia, we only point out the fact that Berg et al. (*16*) investigated second primary carcinomas in index patients with leukemia and noted an increasing frequency of carcinomas of the prostate.

We have already stated that in both sexes the second factor has high loadings in intestine and rectum cancers as well as lung cancer, but it should be noted here, that stomach cancer of males has a moderately large negative loading, and that factor scores of U.S.A. white and nonwhite are negative, despite the fact that this factor has positive correlations with intakes of meat, animal products and oil and fats. Scrutinizing the geographic distribution of the scores of this factor, we note that high loadings were found in relatively northern parts of the world such as Denmark, North Ireland, and England, thus yielding

high negative correlation with the mean atmospheric temperature. Considering the fact that excessive intakes of oil and fats especially animal oil as well as total energy measured in terms of total caloric intake have high correlations with the second factor, a possibility of association between rectum or intestine cancer and the excessive intakes of energy or oil and fats may well be suggested. This hypothesis conforms to the fact that in Denmark, Scotland, and Belgium which indicate high factor scores, intake of oil and fats is more than in the U.S., where high factor score is computed in the first factor. Hence, rectum and intestine cancers are suspected to have a strong association with a high intake of fats.

From the result of factor analysis based on 27 cancer sites of males and females, it is seen in Table 16 that the third factors in males (Table 10) and females (Table 12) are completely heterogenous with respect to the geographical distribution, although esophagus, oral, and larynx cancers are clustered differently in each sex. In males, the correlations shown in Table 14 conform to the well-known hypothesis that excessive intake of alcohol is a main etiological cause of these cancer sites. We also note that alcohol is abundant in France, and oral and esophagus cancers occur more frequently in France.

Next, we shall give some comments on the results of stepwise regression procedure shown in Tables 17 and 18. The result that alcohol was selected first in oral, esophagus, and larynx cancers may well reflect the validity of the hypothesis stated above. Besides alcohol, it is interesting to note that intake of fruit is negatively correlated with the mortalities due to oral, esophagus, and larynx cancers in both sexes, since it is generally believed that excessive drinking and insufficient fruit intake may sometimes cause vitamin deficiency, which is suspected to be a possible etiologic factor for these cancer sites. Meat and rainfall have regression coefficients opposite in sign in case of stomach and intestine cancers, that is, meat has a positive regression coefficient on intestine cancer but a negative regression coefficient on stomach cancer. On the other hand rainfall is in the situation opposite to meat. It is well-known that Japan and Chile where high incidence of stomach cancer is observed, have much rainfall throughout the year, and people in these countries do not eat a great deal of meat at a time. Hence, from our result, it seems valid to postulate that excessive intake of meat is an etiological factor of intestine cancer, as it has been suggested by many researchers. It should be noted, however, that it is inappropriate to conclude that excessive intake of meat would function as inhibitive factor for stomach cancer, since the statistically significant correlation does not always imply a casual relationship between the variables, and high correlation may sometimes be caused by the existence of a third variable. Note that such correlation is termed as spurious correlation, and existence of such a correlation may sometimes lead to incorrect interpretation of the result. For example, let us consider the case of sugar and pancreas cancer in males shown in the bottom row of Table 17. Presumably, it might be better to explain the high correlation between sugar and pancreas cancer by the spurious correlation in terms of a hypothetical variable, say, the degree of industrialization, since sugar is abundant in the industrialized countries and the mortality of pancreas is also high in such countries as well. But we cannot deny the possibility of a casual relationship between sugar and pancreas cancer at all, since in view of clinical epidemiology, an adequate supply of sugar may sometimes work as a remedy for patients suffering from diseases of pancreas. Finally, we should mention that etiological factors such as excessive alcohol drinking and meat intake which this study revealed, are not considered as the initiator but as the promoter in the pathogenesis of cancer.

## Methods of Analysis Used in the Study

Finally, we shall illustrate some points on the methodology of the analysis used in this study. First of all, we should point out that the data used in the analysis are not entirely appropriate in sampling for reasons that most of 24 countries available in our analysis are distributed in the European countries, and sites of cancer such as bladder, kidney, and ovary are not included in the analysis. In view of this fact, we cannot assume that 24 countries used for the analysis are random samples from the population of the world; hence we can hardly say that our results reflect the global tendency in the mutual relationship of the cancers at large. Of course, we cannot deny the possibility that the additional data from the other countries may change the whole implications of the obtained factors in this study, but it is not likely that the obtained results are different from the present ones even if we add data from many other countries.

From the methodological point of view, it is very difficult to estimate the change in the factor structure which would result from addition of new variables as well as new data. We should admit that development of such powerful methodology would make it possible to elucidate the meaning of the results obtained in our study. In addition, we expect that statistical data on cancer deaths in all the countries in the world, including many developing countries for which data are hard to obtain, would become available for the

biostatisticians who are concerned with interrelationship between cancer mortality and various environmental factors.

As for the statistical data, we should mention the reliability of the data obtained from the death certificate, which may sometimes affect the accuracy of the results. It is well known that the reliability of the death certificate data may vary in each geographic area. Studies carried out in the U.S. (*17, 18*) on the accuracy of the death certificate revealed that the most misleading errors in the certificate of the cause of cancer death lies in the geographical differences in the assignment of cause of death to each of the malignant tumors. Such errors are liable to happen especially in the sites such as prostate or bile duct, which physicians find difficult to diagnose. With regards to prostate cancer, Maruchi (*19*) speculated that the reported number of cases of prostate cancer according to death certificates in Japan is estimated to be less than two thirds of the real number by comparing the reported number in the Japanese Archives of Autopsy Findings with that from the death certificate. In view of the evidence, some correction is needed in order to estimate the true figures of deaths from cancer, in which the accuracy of diagnosis is not satisfactory for various reasons.

Finally, we should discuss the usefulness of factor analysis. Some biostatisticians or epidemiologists may raise questions whether factor analysis by itself provides an etiological factor for some diseases. Factor analysis or principal component analysis are said to be methods which reduce the information contained in each variable without losing useful information and which provide clues useful for finding etiological factors. When there are many correlated variables which are supposed to be related to the etiology of a certain disease, univariate statistical analysis such as testing of the differences of mean values carried out separately for each variable may lead to misleading results, since it ignores the correlations among the variables.

Hence, we might say that usefulness of application of factor analysis to epidemiological data lies in finding some hints which may well lead to detecting etiological factors of some diseases hidden behind the data, which could not be detected without application of factor analysis. As an alternative method of finding some clusters of various sites of cancer, Burbank (*20*) recommends cluster analysis, since its application would produce the successive hierarchical structure of various cancer sites. But such hierarchical structure does not necessarily correspond to pathogenesis of cancer in the human body. Furthermore, we must mention that clusters resulting from an application of cluster analysis technique are usually interdependent, and results of cluster analysis never produce information related to the mutual relationship among the extracted clusters. Hence, we would like to emphasize that factor analysis has an advantage over cluster analysis in the usefulness of the clusters.

This does not imply that we totally deny the usefulness of cluster analysis. If we apply both factor analysis and cluster analysis to the same data, it would be a powerful method to develop an etiological hypothesis, provided that we could obtain all the data needed for the analysis of our target. With these points in mind, we would like to develop our study further in order to pursue the unknown etiology of cancer, using the various methods of biostatistics.

We thank Prof. Y. Takane, Department of Psychology, McGill University, Montréal, Canada, for checking the English manuscript. Thanks go to Mrs. M. Yamagami, Department of Epidemiology, Tokyo University, Japan, for providing us with the data on food habits in the world.

## REFERENCES

1. Burkit D. P. Related disease-related cause. Lancet ii: 1229 (1969).
2. Burkit D. P. Relationship as a clue to causation. Lancet ii: 1237 (1970).
3. Segi, M. Cancer Mortality in Japan. Tohoku University, Sendai, Japan, 1965.
4. Segi, M. and Kurihara, M. Age-Adjusted Death Rates and Specific Death Rates from Cancer of Prefectures in Japan, 1958-1967, Segi-Institute of Cancer Epidemiology, Nagoya, Japan, 1972
5. Segi, M. Cancer Mortality for Selected Sites in 24 Countries from No. 1 to No. 10. Tohoku University, Sendai, Japan, 1972.
6. Dunham, L. J., and Bailer, J. C. World maps of cancer rates and frequency ratios. Part 1. Ind. Med. 39: 62 (1970).
7. Hoover, R., Mason, T. J., McKay, F. W., and Fraumeni, J. F. Geographic Patterns of Cancer Mortality in the United States, DHEW, Washington, D.C., 1970.
8. Waterhouse, J. A. H. Cancer Handbook of Epidemiology and Prognosis. Churchill Livingstone, Edinburghs, London, 1974
9. Yanai, H., Inaba, Y., Takagi, H., and Yamamoto, S. An epidemiological study on mortality rates of various cancer sites during 1958 to 1971 by means of factor analysis. Behaviormetrika, 4: 55 (1978).
10. United Nations Organization. Monthly Bull. Agri. Economics Statistics 25: No. 4 and No. 7, United Nations, FAO, 1976.
11. Harman H. Modern Factor Analysis. Univ. Chicago Press, Chicago, 1960.
12. Guttman, L. Multiple rectilinear prediction and the resolution into components. Psychometrika, 5: 75 (1940).
13. Kaiser, H. F. The Varimax criterion for analytic rotation in factor analysis Psychometrika 23:187 (1958).
14. Draper N. R., and Smith, H. Applied Regression Analysis. Wiley New York, 1966.
15. Lee, A. M., and Fraumeni, J. F. Arsenic and respiratory cancer in man: an occupational study. J. Natl. Cancer Inst. 42: 1045 (1969).
16. Berg, J. W., Schottenfeld, D., and Ritter, F. Incidence of

multiple primary cancers. III. Cancers of the respiratory and upper digestive system as multiple primary cancers. J. Natl. Cancer Inst. 44: 263 (1970).
17. U.S. Public Health Service. Comparability of Age on the Death Certificate and Matching Census Record, United States May-August 1960. USPHS Publ. No. 1000, Series 2, No. 29, GPO, Washington, D.C., 1968.
18. U.S. Public Health Service. Comparison of the Classification of Place of Residence on Death Certificates and Matching Census Records, United States, May-August, 1960. USPHS Public. Series 2, No. 30, GPO, Washington, D.C., 1969.
19. Maruchi, N. Epidemiology of cancer of the prostates with special emphasis on U.S.-Japanese comparison. Cancer Clin., 21: 137 (1975).
20. Burbank, F. A sequential space time cluster analysis of cancer mortalities in the United States: etiologic implications. Am. J. Epidemiol. 95: 393 (1972).